Gloss as an aspect of the measurement of appearance

Wei Ji

Department of Colour and Polymer Chemistry, University of Leeds, Leeds, LS2 9JT, United Kingdom

Mike R. Pointer

National Physical Laboratory, Hampton Road, Teddington, TW11 0LW, United Kingdom

Ronnier M. Luo

Department of Colour and Polymer Chemistry, University of Leeds, Leeds, LS2 9JT, United Kingdom

John Dakin

DigiEye PLC, Quartz Close, Warrens Business Park, Enderby, Leicester, LE19 4SG, United Kingdom

The perception and measurement of gloss is just one part of the concept of appearance measurement. We describe the comparison between data derived by 20 observers who each scaled the perceived gloss of a set of 84 neutral and colored test samples and measurements of those samples made using a glossmeter and a sphere-based spectrophotometer. With a glossmeter, the relationship between the visual-scaled data and the measured gloss values could be described by a three-part linear fit or, with a higher correlation, a cubic function. For the sphere-based spectrophotometer, the difference in the luminance factor, ΔY, between specular-included and specular-excluded measurements, was found to give a better linear correlation with the observer-scaled data. The color of the samples did not appear to significantly change the perceived gloss of the surface.   © 2006 Optical Society of America

OCIS codes: 120.0120, 120.5700, 330.5510, 330.1690.

1. INTRODUCTION

Appearance can be described in terms of the ability of an observer to recognize an object, and this recognition is possible because of the interaction between light and the material of the object. The four major things that can happen to light when it encounters an object are

- Specular reflection at the first surface of the object—associated with gloss;
- Absorption within the object—largely responsible for color;
- Scattering within the object—associated with diffuse reflection and transmission: translucency and haze;
- Transmission through the object—associated with transparency, opacity, and clarity.

It has been suggested that the characterization of these optical properties of materials can be divided into at least four groups: color, gloss, translucency, and surface texture.[1] These groups are perhaps not definitive, but they represent useful categories for measurement, especially in view of the fact that measurement techniques already exist for some of them.

Currently, a glossmeter relies on a scale defined according to a physical principle.[2] It is important that the associated psychophysical scale derived from visual data should correlate with the physical scale, but anecdotal evidence has suggested that this may not be the case.

Previous studies[2-4] have been based on the use of neutral-colored samples and the gloss parameter quantified by the use of specialized instruments such as a goniophotometer or a glossmeter. These instruments can measure only one spot on an object at a time and usually assume a fixed illumination and viewing geometry. This is not the same as human visual perception, where the whole object is viewed.

This study will investigate the relationship between the measured gloss of each set of samples and the corresponding perceived value of gloss obtained by a scaling method using a group of observers. The instruments include both a conventional glossmeter and a sphere-based spectrophotometer, and the samples are both neutral and colored.

2. GLOSS

Gloss is the attribute of a surface that causes it to have a shiny or lustrous, metallic appearance and is generally associated with the specular reflection of light from the surface of the object.[3] The amount of this specular reflection can vary from one surface to another depending on several physical parameters:

- The fraction of light reflected in the specular direction.
- The nature and extent of the light that is spread to either side of the specular direction.
- The change in specular reflection as the specular angle changes.

Hunter was one of the first to be aware of these variations in specular reflection, and he designed, in 1934, a glossmeter to compare the capabilities of materials to reflect light at 45° to the surface normal.[2] He studied many different materials and concluded that there were six visual criteria that could be used to rank, and potentially to measure, their gloss:

- Specular gloss—the ratio of the light reflected from a surface at a specified angle to that incident on the surface at the same angle on the opposite side of the surface normal;
- Sheen—the gloss at grazing angles of incidence and viewing;
- Contrast gloss or luster—the ratio of the specularly reflected light to that diffusely reflected normal to the surface;
- Absence of bloom—a measure of the absence of haze or the milky appearance adjacent to the specularly reflected light;
- Distinctness of image—the sharpness of the specularly reflected light;
- Surface uniformity—a measure of the freedom from surface nonuniformities such as texture.

Thus, a surface can appear very shiny if it has a well-defined specular reflectance at the specular angle. The perception of an image reflected in the surface can be degraded by appearing unsharp or by appearing to be of low contrast. The former is characterized by the distinctness of image, and the latter is characterized by the haze or contrast gloss (haze is the inverse of absence of bloom). An added complexity is due to surface nonuniformity leading to an effect known as orange peel. This effect can be caused, for example, by uneven coating of the acrylic overcoat on an automobile finish, leading to a relatively low-frequency ripple.

3. GLOSS MEASUREMENT

As mentioned earlier, the measurement of gloss is based around a physical concept, and the gloss unit is defined such that a glass material of refractive index, n, equal to 1.567 is given a value of 100 GU for any angle of incident light.[2]

Many developments of gloss measurement have been carried out as part of the technical work of the American Society for Testing and Materials (ASTM), commencing in 1925 with an instrument constructed by Pfund.[4] This instrument used parallel light to illuminate the sample at 20° with a detector placed at 20° on the opposite side of the normal. Hunter and Judd later incorporated this design into an ASTM standard method[5] that designated three angles (20°, 60°, and 85°) for measurement, depending on the relative gloss of the surface, and measurements are made relative to the black glass standard.
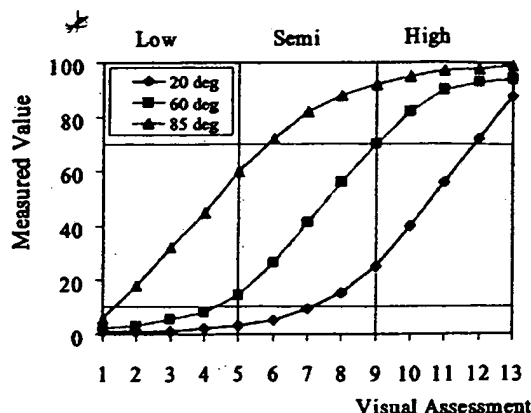

Fig. 1. Measurements of a series of samples, at three angles of illumination, as a function of the rank order of a visual assessment of their gloss.

In order to differentiate the gloss of different samples, it is necessary to select the appropriate measuring geometry. The sample is first measured with 60° geometry. If the gloss value is higher than 70 (high gloss), then it is remeasured at 20°, and, if the gloss value is less than 10 (low gloss), it is remeasured at 85°. These value were loosely defined following some experimental work described by Byk-Gardner[5,6] in which 13 black glass tiles were visually ranked from matt to high gloss and the results compared with measurements at the three angles; see Fig. 1. It was found that the largest difference could be detected by using 85° for low-gloss samples, 60° for semigloss samples, and 20° for high-gloss samples. This appears to be the background to the recommendation that high-gloss, medium gloss, and low-gloss samples should be measured at 20°, 60°, and 85°, respectively.

Standardization in gloss measurement was led by Hunter and ASTM, who first produced ASTM D523 "Standard test method for specular gloss" in 1939. This standard incorporated a method for measuring gloss at a specular angle of 60°. Later editions of the standard (from 1951) included methods for measuring at 20° (high gloss) and 85° (matt, or low, gloss).

A number of instruments are commercially available that conform to the above standard in terms of their measurement geometry, and the standard sets tolerances for the various angles involved.

4. PREVIOUS EXPERIMENTS

The above sections of this paper discuss the establishment of the measurements of gloss and the physical principles on which these measurements are based. Harrison and Poulter,[7] after extensive studies in which the ability of individual observers to rank a range of surfaces according to their gloss was compared with instrumental measurements, have concluded that the gloss of surfaces is not a simple physical property but a psychological gestalt, that is, an appraisal of the physical situation taken as a whole. It was found that observers tend to fall into one of two categories. In the first, observers judge gloss with reference to the total amount of light reflected from a surface, and the others judge gloss according to the sharpness of images seen via the reflecting surface.

A major comprehensive study of visual gloss scaling was made by Billmeyer and O'Donnell.[8] Painted panels were prepared with a wide range of gloss and three achromatic colors (white, gray, and black) by varying the composition and heat treatment of an automobile acrylic lacquer system. Magnitude scaling was used to estimate the visual dissimilarities between members of all possible pairs of specimens in series of 15–20 samples. The data were analyzed using multidimensional scaling techniques. Measurements were also made using a conventional three-angle glossmeter (described below). It was considered conceivable that observers would be able to distinguish at least three types of gloss simultaneously: distinctness-of-image gloss, specular gloss, and reflection haze. The results of the multidimensional scaling analysis, however, showed that observers could not reliably distinguish more than one dimension. This perhaps suggests that there is an inherent property of a surface, related to its gloss, that was scaled by the observers and that is independent of illumination and viewing angle. Put another way, the concepts of distinctness-of-image gloss, specular gloss, and reflection haze are essentially all aspects of the same phenomenon.[9] It was also noted that the visual data obtained in this experiment did not always correlate well with associated measurements made using the conventional glossmeter.

Another experiment that should be mentioned is that which describes a model of surface reflectance used in computer modeling and computer design software in order that the shading of objects on the CRT display matches that of the real-world objects that they are trying to mimic. This aspect of a computer-displayed image is important because the perceived shape of an object is heavily influenced by the variation (and uniformity) of the light reflected from its surface. Ferwerda et al.[10] generated a number of computer images of white, gray, and black spheres using a model of surface reflection.[11] The model allowed the diffuse reflectance, the energy in the specular component, and the spread of the specular component to be varied, thus making it possible to produce a systematic array of spheres. The observers were shown pairs of spheres (378 pairs in total) and asked to judge
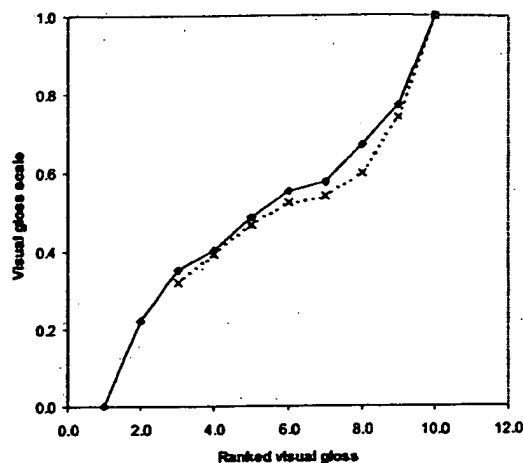
Fig. 3.   Visual scales measured by one observer as a function of ranked sample number of the series. From Obein et al.[15]

their difference in terms of apparent gloss. Multidimensional scaling analysis was then applied; two dimensions were found to adequately describe the data, and these could be related to the apparent contrast of the reflected image (contrast gloss) and the distinctness of the reflected image (distinctness of image). A second experiment was conducted to place metrics on these two axes. Observers were asked to use magnitude estimation to scale the perceptual gloss dimensions, using a subset of the spheres from the first experiment. Analyzing these data showed that the perceived gloss values enabled the derivation of a model that relates the perceived contrast gloss and distinctness-of-image gloss to parameters of a model used to display correctly shaded images on a computer display.[12]

Obein et al.[13–15] argue that gloss is very much a second-order visual attribute in that it results from an interpretation by the brain of first-order signals. This implies that an observer must look at an object from two or three different angles to receive enough information to be able to attribute a value to the gloss of that surface. Their work showed that the scaled visual gloss of a set of samples, obtained using a pair-comparison technique, was not linearly related to the corresponding values obtained from a glossmeter; see Fig. 2. For matt samples, the visual scale undergoes compression, and, for very-high-gloss samples, the gain of the visual responses rises steeply. In the intermediate range, the two scales are almost linearly related. They show that this agrees with the work of Ferwada et al.[10] and Harrison and Poulter.[16] This concept also has some similarities with the findings of Billmeyer and O'Donnell.[8] Analysis of the Obein et al. data also shows that the observers exhibit a form of gloss constancy. When data obtained using two observing angles (60° and 20°) are plotted versus a unique abscissa, for example, the sample number in the series, the two plots superimpose; see Fig. 3. This would indicate that, although the flux that is collected by the eye varies according to the angle of view, an observer is able to recover a visual gloss index that is inherent to the surface. Thus, just as an observer can assign a color to a sample under lights of different spectral power distributions, he can also assign a gloss
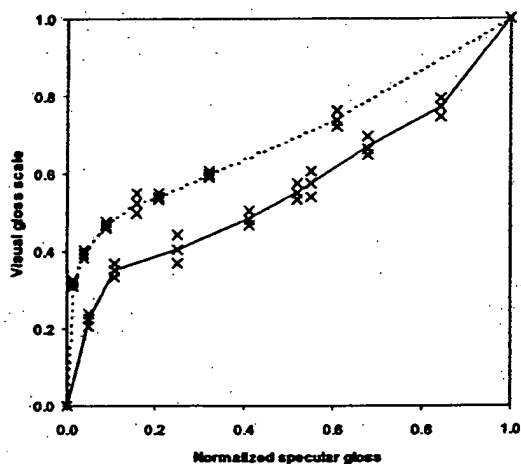

Fig. 2.   Visual scales measured by one observer as a function of normalized specular gloss measured using 60° (solid curve) and 20° (dashed curve) geometry. From Obein et al.[15]

value to a surface despite the change in geometrical distribution of the light, a finding first discussed by Billmeyer and O'Donnell.[8]

5. NEW VISUAL EXPERIMENTS

A. Samples

An experiment was designed in which a number of observers would visually scale the apparent gloss of a series of samples as seen in defined viewing conditions. A set of samples was prepared that consisted of 84 neutral and colored patches. The Swedish Colour Foundation, the manufacturer of the Natural Color System (NCS), kindly donated three neutral gloss scales (white, gray, and black), each with samples at six levels of gloss, and eight NCS colors, each at two levels of gloss. Thus, chromatic group 1 consisted of 16 samples: eight NCS colors at two levels of gloss (about 90 and 17 GU, respectively, measured using a glossmeter with illumination and viewing at 60°). Chromatic group II consisted of an additional 11 NCS colored samples. These were chosen such that they filled in the gaps in color space between the samples in chromatic group I. They were each available at only one, high level of gloss. Two additional levels of gloss samples were produced by hand spraying a gloss-coating agent onto duplicate samples of some of the colored matt samples and hand spraying a matt-coating agent onto duplicate samples of some of the colored gloss samples. The

Table 1. Subsets of the Set of 84 Test Samples

Sample Group	Number of Colors	Gloss Levels for Each Color	Number of Samples
Neutral	3	6	18
Black pile	1	1	1
Chromatic group I	8	2	16
Chromatic group II	11	1	11
Hand-sprayed cards	19	2	38
Total	—	—	84

two spray coating agents used were 101 Gloss Coating and 103 Matt Coating for Artwork manufactured by Esselte Letraset Limited.

A black-pile sample having a measured gloss value of less than 1 GU (0.2 GU measured by the Sheen 160/T at 60°) was also included.

In total, 84 samples were available for use, and the subsets are outlined in Table 1. The spectral reflectance factor of each of these samples was measured using a GretagMacbeth CE7000A sphere-based spectrophotometer (see later). CIELAB colorimetric coordinates were calculated by the instrument software using CIE standard illuminant D65 and the CIE 1964 standard colorimetric observer. The distribution of the samples in a color sense is shown in Fig. 4(a) in a CIELAB a^*b^* diagram and Fig. 4(b) in a CIELAB L^*C^* diagram. It can be seen that the 18 neutral samples fall close to the neutral point (the intersection of the a^* and the b^* axes), and there is a reasonable distribution of colored samples covering a large range of both chroma and lightness.

Each sample was visually examined to ensure that the sprayed coating, and hence the gloss appearance, was uniform over a reasonable area of the surface of each sample. If a sample failed the visual examination, it was resprayed on a new duplicate sample until the uniformity of that sample was acceptable.

The samples were cut to be 3 in. square (7.6 cm by 7.6 cm) and mounted on cards with one extended edge to facilitate handling; see Fig. 5. The thickness of the card was 2 mm to help ensure that the samples remained flat for the duration of the experiment. The color of the card did not affect the appearance of the sample because each gloss sample was opaque.

B. Psychophysical Experiment

The psychophysical experiment was conducted using a Macbeth SpectraLight II viewing cabinet containing illumination that simulated CIE standard illuminant D65, and the samples were assessed against a gray background with a value of CIELAB lightness, L^*, of 70.

In the experiment, observers were given a neutral (gray) reference gloss sample that they were told to assign
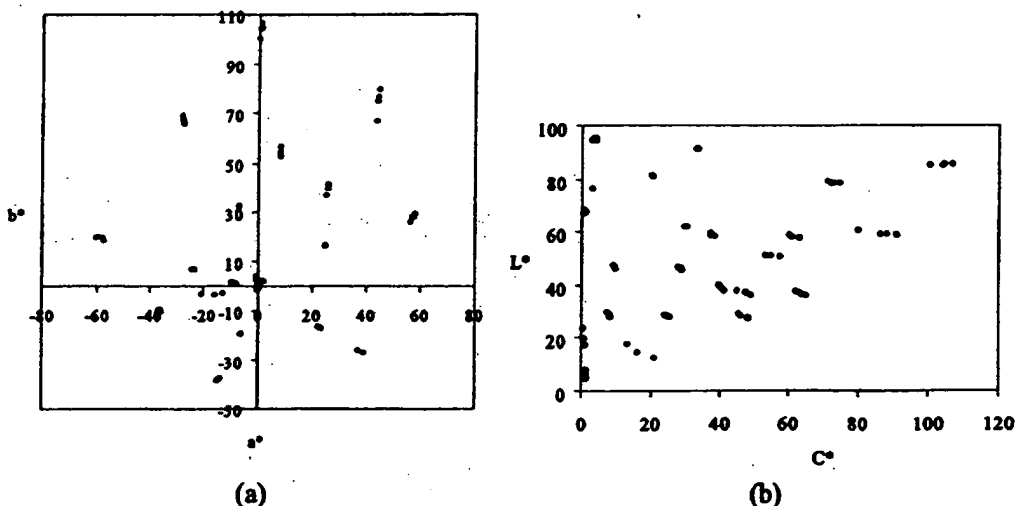

(a)

(b)

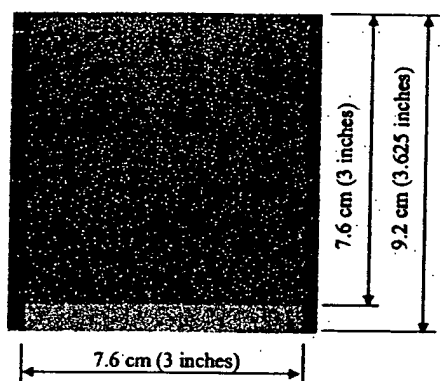Fig. 4. Distribution of the 84 samples in (a) a CIELAB a^*b^* plane and (b) a CIELAB L^*C^* plane.

Fig. 5. Physical size of each sample after it was mounted on a card (the background of the image was a checked board).
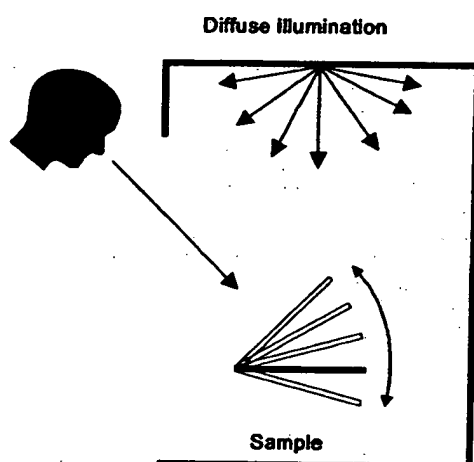
Diffuse Illumination

Fig. 6. Geometric arrangements for the scaling experiment.

a value of 50 GU and asked to scale the gloss of each presented sample on a scale where zero represented no perceived gloss. The observers were not given an upper limit to the scale that they used because the principle of magnitude estimation scaling is that the observers should have an open-ended scale. In practice, however, the observers did tend to use a scale value of close to 100 for the sample that was seen to be of very high gloss. Each sample was compared side by side with that reference sample having an L^* value of 68 and a gloss unit value of 43 at the measurement angle of 60°. The positions of the reference sample and the test sample were interchanged, and the testing samples presented to each observer had a different random order. This scaling method has been widely used in vision experiments and is known as magnitude estimation.

Thirteen observers, who each passed the Ishihara color vision test, scaled 84 test samples, and seven of these observers repeated their observations to give 20 ($2 \times 7 + 6$) sets of observer data. Observers were asked to hold the pair of samples with two hands at arm's length and tilt the test and the reference samples to find an angle of viewing that gave the maximum perceived gloss. This constraint-free assessment method represents the viewing method typically used for assessing gloss in industry; see Fig. 6. To preserve their integrity, the observers were

asked to wear cotton gloves to prevent fingerprints being left on the samples.

Each of the 13 observers finished his/her session in one go, and the length of time taken was between 30 and 45 min. The observers were British (four males), French (one male), Chinese (three males and three females), Korean (one female), and Tai (one female). They were all research staff or research students at the Department of Colour and Polymer Chemistry, University of Leeds, and their age range was from 24 to 45 yr.

A training session was arranged, prior to the main experiment, in which the samples from a Munsell gloss scale,[17] which comprise black samples at five gloss levels, were scaled using the method to be used in the main experiment. These samples were not included in the main set of samples because they duplicated some of the NCS black samples, and the data generated as part of the training exercise were not used in the final analysis.

C. Data Analysis

As mentioned earlier, the data were gathered using the magnitude estimation method. Hence, the data are presented on a ratio scale, and the geometric mean was used as the measure of the central tendency.

Since the individual data are either logarithmic or power functions of the stimulation, they will each be related to the geometric mean function by a power transform.[18–21] This automatically establishes a basis for normalizing the results of individual observers:

$$\log_{10} \bar{S}_i = a \log_{10} S_i + b, \qquad (1)$$

$$S_i' = 10^{\log_{10} \bar{S}_i - b/a}, \qquad (2)$$

where S_i and $\bar{S}_i$ are the individual observer's raw data and the geometric mean calculated from all observers, respectively. The coefficients a and b (the scaling factor and intercept of line described by Eq. (1)) were obtained for each individual observer using the least-squares fitting method between the values of the individual data, $\log_{10}(S_i)$, and the values of the geometric mean, $\log_{10}(\bar{S}_i)$. The values are given in Table 2. These results showed that, although all the values of the coefficients a are close to unity, they still varied by 31%, i.e., from 0.81 to 1.12, in the magnitude of the raw data. For the values of the coefficient b, all results were close to zero. By application of the values of the coefficients a and b to each individual observer's data, those data could be normalized to a common scale.

Note that observer 4 seemed to behave differently from all the other observers. Although the values of the coefficients a and b for this observer are close to unity and zero, respectively, the associated coefficient of variation (CV), as indicated later in Table 4, is large.

Figure 7(a) shows the original observer data plotted against the geometric mean of those data; Fig. 7(b) shows the normalized individual observer's data, S_i', plotted against the same geometric mean of the observer data, $\bar{S}_i$. It can be seen in Fig. 7(b) that the process of normalization serves to reduce the scatter between the two sets of data.

Table 2. Coefficients of Scaling Factor, *a*, and Intercept, *b*, for Each of the 20 Observers

Observer and Coefficient Value				
Scaling factor, *a*				
1	2	3	4	5
0.9582	0.9994	0.8671	0.9593	0.9514
6	7	8	9	10
0.8115	1.0254	0.9051	0.8404	0.9839
11	12	13	14	15
0.9554	1.1184	0.9057	0.9023	0.9834
16	17	18	19	20
1.0921	0.8771	0.8308	0.8173	0.8477
Intercept, *b*				
1	2	3	4	5
0.0194	0.0008	0.1628	0.0441	0.0463
6	7	8	9	10
0.4065	−0.0154	0.2328	0.2695	0.0020
11	12	13	14	15
0.0954	−0.2059	0.1488	0.1869	0.0108
16	17	18	19	20
−0.1880	0.1720	0.2855	0.2960	0.2220

Two methods were used to aid the comparison between any two data sets. The scatter diagram was employed to visualize their difference; and, because in many comparisons the data should be linearly related, the correlation coefficient, *r*, or, more correctly, the coefficient of determination, r^2, was used as a measure of the concordance between the two data sets. The CV was also used as a statistical measure to investigate the agreement between any two sets of data, for example, *x* and *y*. CV is a measure of the distance in the *y* direction of the points from the 45° line in the plot of *y* versus *x*: It expresses the root-mean-square deviation of the distances of the points from this line as a percentage of the mean value of the data set. This statistic has the advantage of giving results that are independent of the size of the data set. It can be thought of as the relative percentage error and is calculated using the following equation:

$$CV = \frac{100}{\bar{y}} \sqrt{\frac{\sum (x_i - y_i)^2}{n}}, \qquad (3)$$

where *n* is the number of samples in the *x* and *y* data sets and $\bar{y}$ is the mean value of the *y* set. A value of CV equal to zero indicates perfect agreement between the two data sets being compared. The CV has a sensible meaning only when it is calculated using similar scales for the *x* and *y* data.

D. Results
Seven of the observers repeated the scaling of the 84 samples. The values of the coefficient of determination and the CV associated with the seven sets are given in Table 3.

The values of r^2 range from 0.9974 to 1.0000 with a mean value of 0.9989. The average value of CV is 3.7, indicating that, on average, the variation between the first and the second sets of data is 3.7%. This result reflects the typical observer repeatability on scaling gloss perception. In an experiment to scale the colorfulness of a series of test colors using the magnitude estimation method, for example, corresponding values of 19% was obtained.[21] In that experiment, the geometric mean was also used for colorfulness to represent the panel results because it was an open-ended scale, i.e., only the zero (neutral) and a reference colorfulness sample assigned a number of 40 were used, similar to the current study. The result in that study shows that scaling colorfulness is much less accurate than scaling gloss.

The lower part of Table 3 shows a similar analysis made using just the 18 neutral samples from the NCS gloss scale. It can be seen that there is not much difference between scaling all samples and scaling only neutral samples.

Another way of understanding the implications of differences in the scaling data of different observers, obtained using the same samples, is to consider the following four cases involving two observers, each scaling a series of 18 samples that progressively increase in gloss.
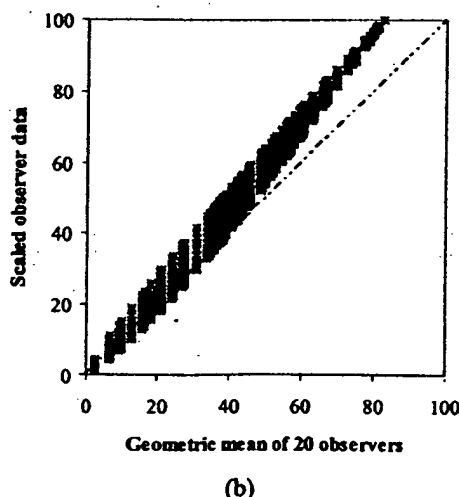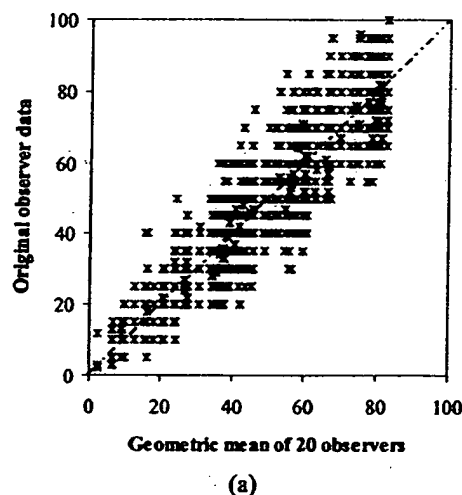

(a)

(b)

Fig. 7. (a) Original observer-scaled gloss data and (b) the scaled gloss data for each of the 20 observers plotted against the geometric mean of those 20 sets of data.

Table 3. Values of the Coefficients of Determination, r^2, and Variation, CV, for the Seven Observers Who Repeated the Scaling of the 84 Samples[a]

	Observer and Coefficient Value							
All samples (84)								
	1	2	3	4	5	6	7	Mean
r^2	0.9984	0.9974	0.9977	1.000	0.9990	0.9997	0.9999	0.9989
CV	4.2	5.4	5.0	0.1	3.3	2.0	1.3	3.1
Neutral samples (18)								
r^2	0.9989	0.9980	0.9984	1.0000	0.9993	0.9998	0.9999	0.9992
CV	5.1	6.4	6.2	0.2	4.1	2.3	1.5	3.7

[a]Data are presented for all 84 samples and for the 18 neutral samples.

Table 4. Values of the Coefficients of Determination, r^2, and Variation, CV, for the 20 Data Sets Compared with the Geometric Mean Data Set[a]

	Observer and Coefficient Value						
	1	2	3	4	5	6	7
r^2	0.9999	1.0000	0.9981	0.9958	0.9998	0.9957	0.9999
CV	1.1	0.3	4.6	6.9	1.4	7.0	1.1
				Observer			
	8	9	10	11	12	13	14
r^2	0.9992	0.9971	1.000	0.9999	0.9987	0.9992	0.9991
CV	3.1	5.7	0.3	1.2	3.9	3.1	3.2
				Observer			
	15	16	17	18	19	20	Mean
r^2	1.0000	0.9991	0.9984	0.9967	0.996	0.9974	0.9894
CV	0.3	3.1	4.2	6.1	6.7	5.4	3.4

[a]All 84 samples were used.

1. Observer A scales the samples using a scale with values from 5 to 90 in steps of 5. Observer B scales the same samples, giving the same numbers except for one, which is scaled 5 units higher (or lower). The associated values of the coefficient of determination and CV are close to 1.00 (actually 0.9981) and 2.5%, respectively.

2. Observer A scales the samples as before. Observer B scales the same samples, giving numbers that are consistently 5 units higher. The associated values are 1.00 and 9.5%, respectively.

3. Observer A scales the samples as before. Observer B scales the same samples, giving numbers that are alternately 5 units higher and 5 units lower. The associated values of the coefficients are 0.96 and 10.5%, respectively.

4. Observer A scales the samples as before. Observer B scales the same samples, giving numbers that are equally spaced by 4 units, thus $5 \equiv 5$, $17 \equiv 20$, $41 \equiv 50$, $73 \equiv 90$. The associated values of the coefficients are 1.00 and 25.5%, respectively.

Table 4 shows the values of the coefficient of determination and CV for each individual data set tested against the geometric mean value of the 20 data sets.

The values of r^2 range from 0.9957 to 1.0000 with a mean value of 0.9894, which is commendably high. The average value of the CV is 3.4. The range of the values is from 0.3% to 7.0%. Again, these values are small.
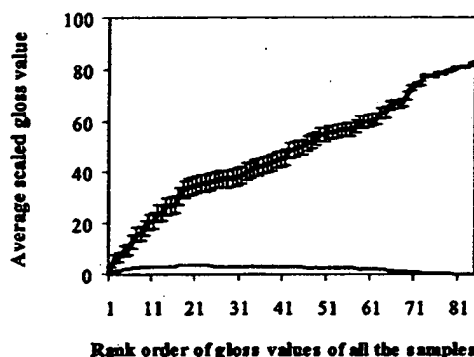

Fig. 8. Average scaled gloss value, for all 84 samples, in rank order of gloss values of all the samples. The error bars represent the associated value of ±1 standard deviation for each sample. The values of the standard deviation are also plotted at the bottom of the graph as a function of the rank order of gloss value.

Figure 8 shows the geometric mean value of gloss for all 84 samples plotted in rank order of gloss values of all 84 gloss samples. The error bars represent the associated value of ±1 standard deviation. The values of the standard deviation are also plotted at the bottom of the graph as a function of scaled gloss value. The latter plot shows that, after an initial rise in value of the standard deviation, the value remains approximately constant within a

defined range until a certain amount of gloss value. The low values of standard deviation are associated with samples with very low gloss and very high gloss, whereas the middle range of scaled gloss value is likely to be consistent.

Thus, a series of 84 gloss samples have been scaled by 20 observers, a mean value obtained for each sample, with an associated value of standard deviation.

6. MEASUREMENTS

Each of the 84 samples was measured using two instruments: a glossmeter and a sphere-based spectrophotometer.

A. Glossmeter
The Sheen Tri-Gloss 160/T glossmeter[22] provides three readings of a sample using illumination angles of 20°, 60°, and 85°. Its configuration is similar to the generic instrument described above.[5]

B. Spectrophotometer
The second instrument that was used to measure the samples was a GretagMacbeth CE7000A sphere-based spectrophotometer. This instrument uses diffuse illumination into the integrating sphere and detects the light reflected by the sample at 8° to a normal to its surface, designated $d/8°$ geometry. An attachment to the sphere has the facility either to include the light reflected by the sample at the specular angle by a plate fitted to the sphere, which is coated with a material similar to that used to coat the inside of the sphere, or to exclude this light using a so-called gloss trap. Two sets of measurements were made, the first without using the gloss-trap, specular-included (SPIN) configuration and the second with the gloss-trap-fitted, specular-excluded (SPEX) configuration. CIE X,Y,Z tristimulus values were calculated from the spectral reflectance data and the difference, ΔY, calculated between the value of the luminance factor calculated from the data that included the specular component, Y_{in}, and that calculated from the data that excluded the specular component, Y_{ex}. This should provide a measure of the light that is reflected at the specular angle. It is unlikely, however, that the geometry of this aperture complies with the specification prescribed for measuring gloss.[4] Thus, three sets of measurements of the 84 samples were available.

Table 5. Sample Uniformity for the 84 Samples Expressed in Terms of the Coefficient of Variation

Sample Group	Number of Samples	Coefficient of Variation	
		Mean	Maximum
Original samples	45	2.8	8.3
Sprayed samples	38	8.4	15.3
Black-pile sample	1	2.0	—
All samples	84	6.0	15.3

Table 6. Sample Stability in Terms of the Correlation Coefficient, r, and the Coefficient of Variation, CV[a]

	20°	60°	85°
Set 1 versus set 2			
r	1.00	1.00	1.00
CV	2.3	0.6	0.3
Set 1 versus set 3			
r	1.00	1.00	1.00
CV	4.8	3.5	5.1

[a]Measured using the Sheen glossmeter.

C. Sample Uniformity
Some investigatory measurements were made on the uniformity of the samples, and their spatial variation in gloss value is summarized in Table 5.

Each sample was measured in five positions using the Sheen Tri-Microgloss 160/T glossmeter: top right, top left, bottom right, bottom left, and center (designated the five-point averaging method). CVs were calculated between the five measurements and their mean and over the samples in each of the four groups described in Table 5. As might be expected, the value of the CV is larger for the sprayed samples than for the original (unsprayed) samples because the spraying process, which was carried out by hand, contributed to the nonuniformity. All samples were considered acceptable by visual inspection, however, because the variation in uniformity was not visible at the optimum viewing angle for gloss detection. Indeed, the larger nonuniformity in appearance terms was that due to the gloss itself rather than the surface variability.

D. Sample Stability
Measurements were also made using the five-point averaging method to monitor the sample stability during the running of the experiment. Each of the 84 samples was measured twice prior to any observer making any observations and a third time after an interval of 1 month during which time the observations were made. They are called set 1, set 2, and set 3. Table 6 shows the values of the correlation coefficient and the CV, between these three sets of data.

Thus, there was a small amount of variation in the measurements over a short period of time, 0.6% at 60°, with larger variation over the longer time interval during which the experiment was carried out, 3.5% at 60°. It should be noted that the sample stability over time is better than the sample uniformity.

7. DATA ANALYSIS

A. Comparison between Gloss Values and Differences in the Luminance Factor
Figure 9 shows the difference in the luminance factor, ΔY, plotted as a function of the gloss value measured using the Sheen glossmeter. Data for all 84 samples are plotted. It can be seen that the relationship is not linear in that there appears to be a breakpoint in the region of 20–30 GU. The curve shown plotted on the figure com-
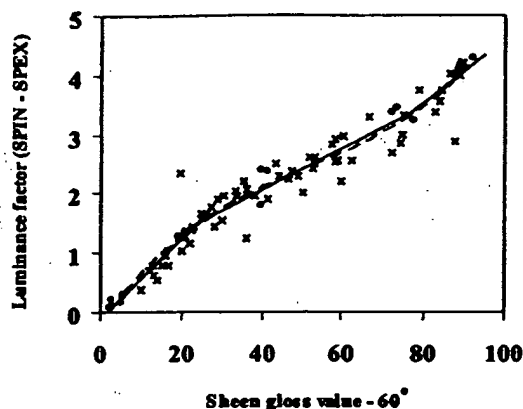
Fig. 9.   Parameter, ΔY, derived from measurements made using a spectrophotometer, as a function of measurements made using a conventional glossmeter with 60° geometry. Solid curve: three-part, linear, least-squares fit. Dashed curve: cubic fit. Crosses and heavy dots represent color and neutral samples, respectively.

Table 7. Statistics of Least-Squares Lines Relating the Measurements Made Using a Glossmeter and a Spectrophotometer[a]

Gloss	n	r^2	m	c
18 neutral samples				
≤20	8	0.9765	0.0690	−0.0926
≥20	10	0.9608	0.0392	0.5573
66 colored samples				
≤20	11	0.5576	0.0803	0.2839
≥20	55	0.9039	0.0365	0.5816
All samples				
≤20	19	0.8527	0.0653	−0.0864
≥20	65	0.8639	0.0382	0.4967

[a]All 84 samples and just the 18 neutral samples were considered.

prises two least-squares curves that have been calculated to give optimum values of the coefficient of determination, r^2, resulting in a breakpoint at a gloss value of 21. The values of the associated statistics are shown in Table 7.

It should be noted that there are only a relatively small number of colored samples with a gloss value less than ~20. An alternative way of considering the trend shown by the data is to fit a cubic function, also shown in Fig. 9. This yields values of the coefficient of determination of 0.9507, 0.9914, and 0.9290 for all the samples, the neutrals, and the colors, respectively. This point will be addressed in a later section of this paper.

B. Comparison between the Visually Scaled Data and the Measured Glossmeter Data

Figure 10 shows the mean of the observer-scaled gloss data, for all 84 samples, plotted as a function of the gloss values measured using the Sheen glossmeter. Again, it appears that there is a breakpoint in the data at a gloss value of approximately 20 units and possibly another at a value of approximately 80 units. These data show the same form as that demonstrated by Obein et al.[15] and described above. To facilitate comparison with their results, we normalized the data plotted in Fig. 10 such that the

maximum value of each scale was unity, see Fig. 11, and fitted least-squares curves to the three separate sections of the data. The values of the coefficient of determination were optimized for the three sections; the resulting curves are also shown in Fig. 11, and their statistics are shown in Table 8.

It can be seen that, though the overall trends in the two sets of data are similar, the values of the slope and intercept of the fitting curves are different. As before, the data can also be fitted with a cubic function giving a coefficient of determination of 0.9746 with a corresponding value of 0.9751 for the Obein et al.[15] data.

C. Comparison between Visual Scale Data and Differences in the Luminance Factor

Figure 12 shows the visually scaled data plotted as a function of ΔY, the difference in the luminance factor for
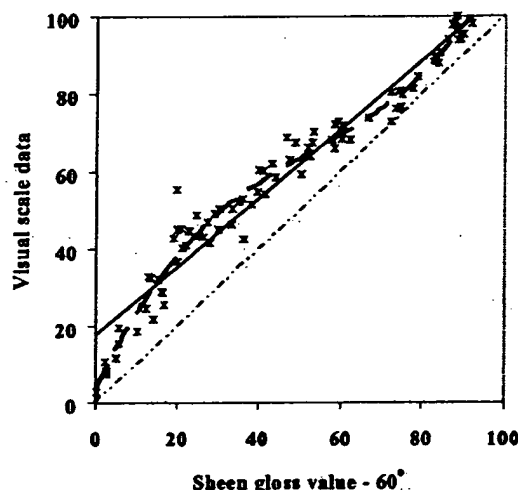

Fig. 10.   Visual scale data as a function of measurements made using a conventional glossmeter with 60° geometry. Solid line: linear, least-squares fit. Dashed curve: cubic fit. (Dotted–dashed line is 45° line).
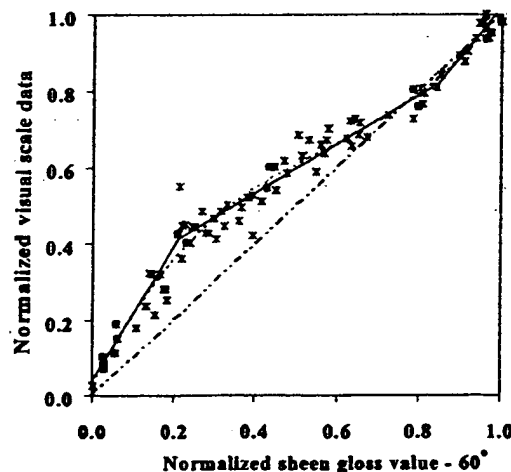

Fig. 11.   Normalized visual scale data as a function of normalized measurements made using a conventional glossmeter with 60° geometry. Solid curve: three-part, linear, least-squares fit. Dashed curve: cubic fit. (Dotted–dashed line is 45° line). Crosses and heavy dots represent color and neutral samples, respectively.

SPIN and SPEX measurements. As might be implied from the comparisons presented above, there is an approximate linear relationship between these two sets of data. A least-squares fit line is also shown, and the associated statistics are given in Table 9.

Also shown in Table 9 are the statistics associated with the same data but after normalization to give a maximum value of unity for each set. The procedure enables the CV

Table 8. Statistics of Least-Squares Lines Relating the Normalized Visual Data to the Measurements Made Using a Glossmeter, Considering All 84 Samples, and Comparison Statistics for the Data of Obein et al.[15]

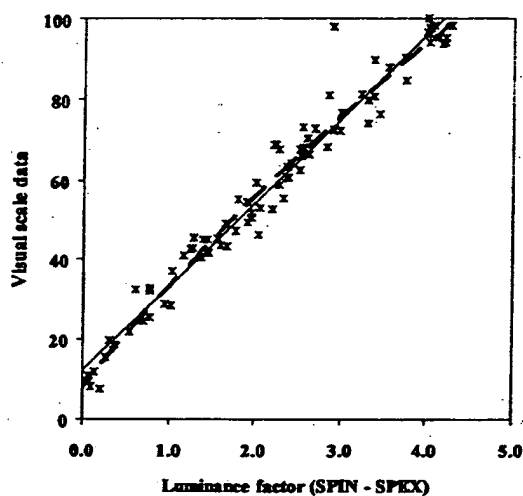| Gloss	n	r^2	m	c
All samples				
≤0.20	25	0.8402	1.4210	0.0508
≥0.20≤0.85	36	0.9051	0.6291	0.2861
≥0.85	23	0.7076	1.0131	−0.0156
Obein et al.[15]				
≤0.20	3	0.9788	3.1571	0.0229
≥0.20≤0.85	7	0.938	0.5842	0.2633
≥0.85	2	1.0000	1.4344	−0.4344

Fig. 12. Normalized visual scale data as a function of a parameter derived from measurements made using a spectrophotometer. Solid line: linear, least-squares fit. Dashed curve: cubic fit.

to be calculated because the normalization process brings both of the measures to the same scale. Applying a cubic to the data shown plotted in Fig. 12 improves the value of the coefficient of determination only marginally to 0.9634 when applied to the total data set of 84 samples, 0.9888 for the neutral samples, and 0.9484 for the colored samples.

8. DISCUSSION

From the above comparisons, it can be concluded that the two glossmeters are giving broadly similar results for measurements at 60° and 20°. The measure derived from spectrophotometric measurements is not linearly related to the gloss measurements and can be considered to have a breakpoint or, perhaps more likely, to be described as a cubic relationship. Similarly, the visual scale data are related to the measured gloss values by a combined linear function or a cubic function. This suggests, and it is subsequently found to be the case, that the visual scale data should be related to the luminance factor difference by a linear function.

Although the visual scale data are linearly related to the measured gloss values for samples of moderate to high gloss, values above ~20 GU, the gloss of samples with lower measured gloss is perceived by the observers to decrease at a faster rate than the measurements would imply. There is some spread in the data that is described by the statistics presented above; the trends suggested by the least-squares- and cubic curve-fitting routines that have been applied to the data all have reasonably high values of coefficient of determination and, where relevant, coefficient of determination that is considered robust.

The finding that the visual scale data are related to measured data via a combination of linear functions, or a cubic function, is not new. Obein et al.[15] have demonstrated the former, and Billmeyer and O'Donnell[8] have shown the latter. Ferwerda et al.[10] also describe data that might be better fitted by a combination of linear functions rather than the single linear function they suggest.

The question arises as to why the better correlation is found for a measurement based on an integrating sphere as opposed to an instrument that uses specular illumination and detection. A possible answer may be that the illumination using the integrating sphere is diffuse and, while the instrument measures that light that is reflected at the specular angle, it is the only light that is diffusely

Table 9. Statistics of Least-Squares Lines Relating the Visual Data to the Measurements Made Using a Spectrophotometer[a]

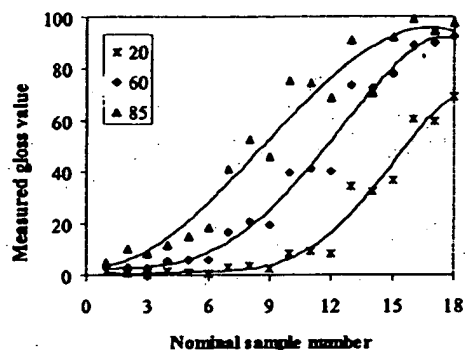| Data	n	r^2	m	c	CV%
Measured and scaled data					
Neutral samples	18	0.9864	20.8503	9.9003	—
Colored samples	66	0.9459	20.7561	13.4861	—
All samples	84	0.9585	20.9439	12.3455	—
Normalized data					
Neutral samples	18	0.9865	0.8965	0.0948	16.3
Colored samples	66	0.9459	0.8938	0.1296	17.8
All samples	84	0.9585	0.9013	0.1186	17.6

[a]All 84 samples were considered.

Fig. 13. Measurements of a series of samples, at three angles of illumination, as a function of the rank order of a visual assessment of their gloss. Similar to Fig. 1 but based on the method described in the paper.

illuminating the sample. This is likely to be much less than that considered in the measurement made using the conventional glossmeter.

In the opening part of this paper, a rationale was suggested as to why different angles of illumination should be used according to the level of the gloss. Figure 1 shows a set of data that attempt to support the argument that the sample should first be measured with 60° geometry and then at 20° if the gloss value is higher than 70 or at 85° if less than 10. It is of interest to plot the measured gloss data by the Sheen glossmeter on 18 neutral samples accumulated in this work as a function of the nominal sample numbers (rank order) that relate to the gloss of the samples; see Fig. 13. It can be seen that this graph shows a passing resemblance to Fig. 1; the points are fitted with cubic functions to illustrate the trend of the data. The spacing between the three data sets is, however, less than that implied by Fig. 1, and this suggests that the argument for using different angles for measurements at different levels of gloss is not strongly supported.

9. CONCLUSIONS

The following have been demonstrated in this work:

• Measurements made using a spectrophotometer, and the subsequent calculation of the difference between the luminance factors for specular-included and specular-excluded detection, do not correlate with the conventional glossmeter measurements over the entire range. At relatively low values of gloss (≤ 20 units), the rate of change falls at a higher rate.

• Visually scaled gloss data are also not correlated with measured gloss data over the entire range and show a sharper fall at low values of gloss. A cubic function better fits the data than three linear lines changing gradient at two breakpoints. This result is consistent with that found by other workers.

• Visually scaled gloss data are correlated with the measure derived from measurements made using a spectrophotometer. In this instance, a cubic fit does not give a significantly higher correlation.

• No appreciable difference was found between the visual data derived using neutral samples and that derived using colored samples. Equally, some of the samples were

constructed by spraying colored papers to provide a range of gloss values, and no appreciable difference was found between the visual data derived using unsprayed samples and sprayed samples.

REFERENCES

1. J. B. Hutchings, *Food Color and Appearance*, 2nd ed. (Aspen, 1999).
2. M. E. Nadal and E. A. Thompson, "New Primary standards for specular gloss measurements," J. Coat. Technol. **72**, 61–66 (2000).
3. R. S. Hunter and R. W. Harold, *The Measurement of Appearance*, 2nd ed. (Wiley, 1987).
4. A. H. Pfund, "The measurement of gloss," J. Opt. Soc. Am. **20**, 23–26 (1930).
5. ASTM D523, "Standard test method for specular gloss," (American Society for Testing and Materials, Gaithersburg, Md., 1999).
6. Byk-Gardner, Technical Note. See http://www.bykgardner.com (2004).
7. V. G. W. Harrison and S. R. C. Poulter, "Gloss measurement of high gloss papers," Pap. Res. 7, 128–136 (1954).
8. F. W. Billmeyer and F. X. D. O'Donnell, "Visual gloss scaling and multidimensional analysis of painted specimens," Color Res. Appl. **12**, 315–326 (1987).
9. R. Sève, "Problems associated with the concept of gloss," Color Res. Appl. **18**, 241–252 (1993).
10. J. A. Ferwerda, F. Pellacini, and D. P. Greenberg, "A psychophysically-based model of surface gloss perception," in Proc. SPIE **4299**, 291–301 (2001).
11. G. J. Ward, "Measuring and modeling anisotropic reflection," in ACM Comput. Graphics **26**(2), 265–272 (1992) [*Proceedings of SIGGRAPH '92*, Chicago, Illinois, July 26–31, 1992 (www.SIGGRAPH.org)].
12. Y. Sun, "A physically-based reflection model for glossy appearance," in *Proceedings of Tenth IS&T/SID Color Imaging Conference: Color Science, Systems and Applications* (Society for Imaging Science and Technology, 2002), pp. 172–177.
13. G. Obein, T. Leroux, and F. Vienot, "Bi-directional reflectance distribution factor and gloss scales," in Proc. SPIE **4299**, 279–290 (2001).
14. G. Obein, K. Knoblauch, A. Chrisment, and F. Vienot, "Perceptual scaling of the gloss of a one-dimensional series of painted black samples," Perception **31**, 65 (2002).
15. G. Obein, T. Leroux, K. Knoblauch, and F. Vienot, "Visually relevant gloss parameters," in *Proceedings of the 11th International Metrology Conference*, Toulon, France, October 20–24, 2003 (www.cfmetrologie.com).
16. V. G. W. Harrison and S. R. C. Poulter, "Gloss measurement of papers—the effect of luminance factor," Br. J. Appl. Phys. **2**, 92–97 (1951).

17. See GretagMacbeth website: http:// www.gretagmacbeth.com (2004).
18. S. S. Stevens, "On the psychophysical law," Psychol. Rev. **64**, 153–181 (1957).
19. S. S. Stevens and E. H. Galanter, "Ratio scales and category scales for a dozen perceptual continua," J. Exp. Psychol. **54**, 377–341 (1957).
20. C. J. Bartleson, "Changes in colour appearance with variations in chromatic adaptation," Color Res. Appl. **4**, 119–137 (1979).
21. M. R. Luo, A. A. Clarke, P. A. Rhodes, S. A. R. Scrivener, and C. J. Tait, "Quantifying colour appearance. Part 1. LUTCHI color appearance data," Color Res. Appl. **16**, 165–180 (1991).
22. See www.sheeninstruments.com (2004).

THE MEASUREMENT OF GLOSS

By A. H. PFUND

SUMMARY

It is proposed to measure the specular-reflection from various materials and to determine objective gloss which defines surface characteristics only. Samples are mounted on a circular cylinder which is spun rapidly. Large angles of incidence are employed. The brightness of the image of a source reflected specularly from the cylindrical sample is photometered by means of an optical pyrometer. Actual results show that gloss so measured is independent of the brightness of the back-ground (diffuse reflection).

When light falls on materials such as enamel, silk, paper etc., specular as well as diffuse reflection takes place. The smoother the surface,the greater is the intensity of the specularly reflected light and the greater, hence, is the lustre or gloss. It is recognized that if two materials, one white and the other black, have the same surface characteristics, the black one will *appear* the glossier. From this statement it is evident that gloss may be defined in two ways: first, according to the sensation of glossiness (subjective) or, secondly, according to the actual surface condition (objective). Subjective gloss is defined[1] as the ratio of the specular to the diffuse reflection. Obviously, if a black and a white material were given the same subjective gloss, the surface characteristics would be violently different. Now, the manufacturers of varnish, paper, fabric, etc., are interested primarily in the surface itself—regardless of the nature of underlying color. In other words, they are interested in producing materials of definite finish regardless of whether the background be green, red or blue. In the following, this problem is attacked by formulating a definition of objective gloss and by describing an instrument which measures the desired quantity.

Objective gloss is here defined as the ratio of the intensity of light reflected specularly from the material in question, at a stated angle of incidence, *to* the intensity of light reflected specularly from a perfectly polished surface of the same material, at the same angle of incidence.

To measure the specular reflection only, it is necessary to suppress the diffuse reflection. This is accomplished by giving to the test-object a cylindrical form and by resorting to large angles of incidence. Because of the cylindrical form, a virtual image of the light-source is formed. For a perfect reflector, and to a greater or lesser extent, for partial reflectors, the intrinsic luminosity of the image of a finite source does not vary as the distance from the source is increased. Consequently, it is

---

[1] L. A. Jones, J.O.S.A., & R.S.I., *6*, p. 147; 1922.

possible to remove the source to such a distance that the diffusely reflected light becomes negligible in comparison with the intrinsic luminosity of the image formed by specular reflection. Again, by increasing the angle of incidence, the diffuse reflection decreases while the specular increases. Experience has shown that at an angle of incidence (and of reflection) of 75° or thereabouts, the diffuse reflection has been removed to all intents and purposes for all but the least glossy surfaces. For surfaces having very high gloss it has been found best to use nearly normal incidence (5°) while for surfaces which are very matte, grazing incidence (85° or more) has been found to yield sharpest discrimination. Considering the wide variety of materials whose gloss is to be measured, it is obviously out of the question to find a perfect reflector for each substance. It is to be noted, however, that paper, fabric, paints, etc., have approximately the same refractive index as glass. Therefore, a cylinder of soda glass blackened on the inner surface, and having a high optical polish, was chosen as the standard, Incidentally it may be mentioned that the cylindrical form given the sample holder has the great advantage of making it easy to give a regular geometrical form to the sample. To prepare a surface that is sufficiently plane for reflection measurements is at times difficult. Anticipating the results to be presented it may be said that the entire range of objective gloss —extending from that of polished glass to that of surfaces fumed with MgO or soot may be covered.

The essentials of the gloss-meter are shown in Fig. 1. Here A is a substantial vertical post bearing a fixed pivot, D, and a movable collar, E. Attached to these is an adjustable parallelogram B $B'C$ C' pivoted at all corners. Illumination of the cylindrical sample H is effected by an incandescent lamp with frosted bulb F. The pivot D is on the same level with the highest portion of the cylindrical sample so that, if the angle of incidence be changed by moving E, the beam reflected specularly from the sample will always lie in the direction of the arm C. The sample-holder consists of a highly polished brass sleeve H which is slid on the solid cylinder G. The latter is contiguous with a shaft mounted in a long bearing I. By means of a small motor and pulley (not shown) the sample is rotated so rapidly that all surface irregularities of the sample are wiped out. If paper or fabric is to be tested, the sample is simply wound around H—the free edge being either glued down with library paste or held down with rubber bands applied at the extremities of the cylinder. Threads of silk, rayon etc. are wound around the cylinder to form a "spool" while paints or varnishes may be
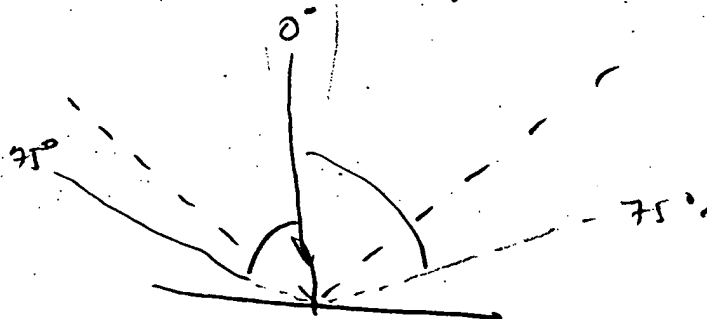
applied to the cylinder directly or better, to a sheet of thin aluminum which is wrapped subsequently about the cylinder. To measure the intrinsic luminosity of the image of F formed by reflection from the cylindrical surface, a small optical pyrometer K, supplied with a red filter R, is used. The specular image is of elongated form and is brightest in the middle. By adjusting the position of the pyrometer-lamp filament so that it crosses the specular image transversely it is possible to bring about the disappearance of the filament in the brightest portion of the image. Readings are taken on the milliammeter connected in series with the pyrometer lamp, battery and rheostat.
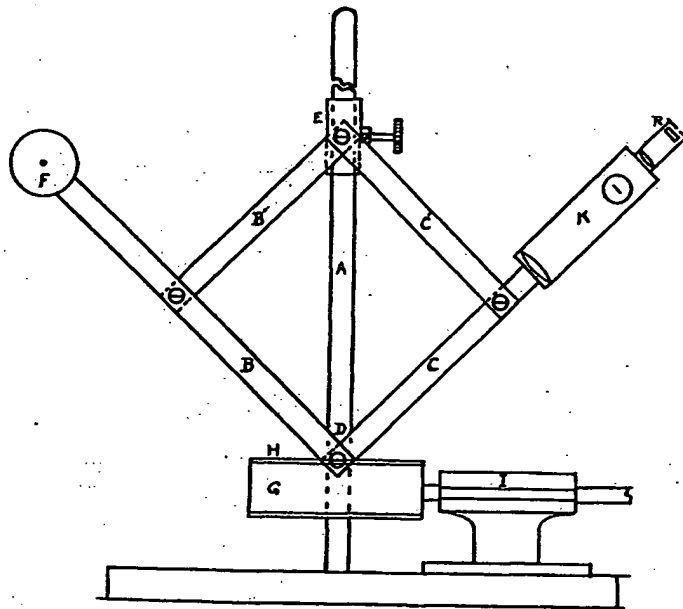

FIG. 1. *Sketch of gloss-meter.*

Calibration of the instrument is effected by mounting the glass standard in position and by bringing about a match in the optical pyrometer. After recording the milliammeter reading, a rotating sector of variable aperture or, a series of neutral tint screens are introduced so as to reduce the intensity of the reflected beam in known ratios. Relative light intensities are finally plotted against milliammeter readings.

The performance of the glossmeter is illustrated by the results shown in the following table. To test the degree of immunity from diffuse

reflection, samples of developing paper were tried out. Each sheet was exposed over half of its area and was then developed. In this manner, identical surface conditions with respective black and white backgrounds were realized.

Material	Gloss	Remarks
1. Glass	1.00	Standard—All angles of incidence 75%
2. This journal J.O.S.A. (print)	0.18	
3. This journal J.O.S.A. (cover 1929)	0.047	
4. Phys. Rev. print	0.07	
5. Bromide paper (black)	0.20	
Bromide paper (white)	0.20	Same finish on both samples
6. News print	0.057	
7. Black Duco	0.17	
8. Glossy Velox (black)	0.78	
Glossy Velox (white)	0.78	Same finish on both samples

While the absolute values here presented are of no particular importance, it is shown however that a wide range of gloss may be covered and, what is perhaps most important, that the value of gloss is determined solely by the surface characteristics and is uninfluenced by the brightness of the background.

JOHNS HOPKINS UNIVERSITY,
   BALTIMORE, MARYLAND.
     OCTOBER, 1929.